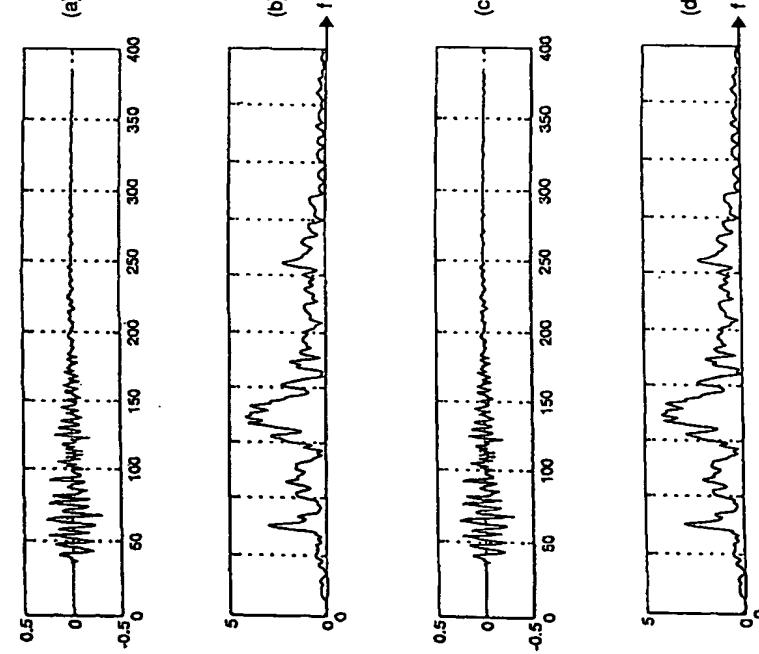




(11) EP 0 948 237 A2  
 (13) Veröffentlichungstag:  
 06.10.1999 Patentblatt 1999/40  
 (21) Anmeldernummer: 89106123.5  
 (22) Anmeldetag: 01.04.1999



Szene Z4

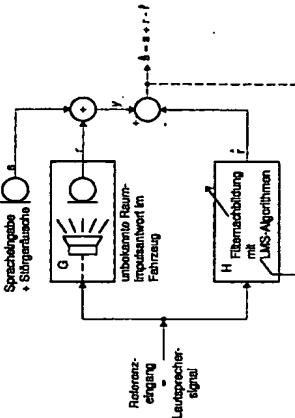


FIG.1

EP 0 948 237 A2

FIG.28

## Beschreibung

[0001] Die Erfindung betrifft ein Verfahren zur Störerfreuung eines Mikrophonsignals, bestehend aus mehreren Verfahren, die eine Verarbeitung des Störungssignals im Frequenzbereich vornimmt, insbesondere für die Sprachausgabe von Kommandos und/oder für Freisprechfunktionen zunehmend an Bedeutung, wobei insbesondere die Situation in einem Fahrzeug ein wichtiger Anwendungsbereich ist.

[0003] Eine besondere Situation ist in Fahrzeugen häufig dadurch gegeben, daß ein Wiedergabegerät wie z.B. ein Radio, ein Kassettent- oder CD-Abspieler über einen Lautsprecher eine Geräuschauslösung erzeugt, die als Störignal ein von einem Mikrofon aufgenommenes Sprachsignal, beispielsweise für den Sprachberichterstatter, ein. Derartige Verhältnisse sind in einem Sprachberichterstatter oder zur verständlichen Sprachübertragung über Telefon ist das Mikrophonsignal sowie wie möglich von Störignalanteilen zu befreien.

[0004] Das von einer Störquelle, insbesondere einem Lautsprecher ausgehende Störignal gelangt nicht nur auf direktem kürzestem Weg zum Mikrofon, sondern tritt auch noch über zahlreiche Reflexionen als eine Übertragung einer Mehrzahl von Echoen mit verschiedenen Lautzeiten im Mikrophonsignal in Erscheinung. Die gesamte Einwirkung des Störsignals von der Störquelle auf das Mikrophonsignal kann durch eine a priori unbekannte Übertragungsfunktion und Filterung des Fahrzeugspeisesignals des Fahrzeugsraumes, eines Kraftfahrzeugs beschrieben werden. Die Übertragungsfunktion ändert sich je nach Besetzung des Fahrzeugs und nach Position der einzelnen Personen. Durch Nachbildung dieser Übertragungsfunktion und Filterung eines Referenzsignals von der Störquelle mit dieser Nachbildung kann ein Kompressionsignal erzeugt werden, welches durch Subtraktion vom Mikrophonsignal ein von Störignal befreites Signal, beispielsweise ein reines Sprachsignal liefert. Im Realfall stellt die genannte Nachbildung eine mehr oder minder gute Annäherung an die unbekannte Übertragungsfunktion dar und die Störung kann nicht vollständig beseitigt werden.

[0005] Aufgabe der vorliegenden Erfindung ist es, ein Verfahren zur Störberichtigung eines Mikrophonsignals anzugeben, das bei verteiltem Signalverarbeitungsaufwand gute Eigenschaften hinsichtlich der Erstellung aufweist.

[0006] Die Erfindung ist im Patentanspruch 1 beschrieben. Die untersuchte enthaltene Verarbeitungsausgestaltungen und Weiterbildungen der Erfindung. [0007] Wesentlich an dem erfindungsgemäßen Verfahren ist, daß die Kompression des Störignalanteils im Mikrophonsignal mittels eines dem Referenzsignal über die Nachbildung der Übertragungsfunktion erzeugten Kompressionssignals im Frequenzbereich vorgenommen wird, so daß ein Mikrophonsignal Kompressionsignal und Ausgangssignal im Frequenzbereich, d.h. in Form von Spektren vorliegen. Die Signaverar-

beitung in diesem Verfahrensschritt im Frequenzbereich erfordert zwar eine spezielle Transformation des Mikrophonsignals, bedeutsam ist aber, daß die Nachbildung der Übertragungsfunktion im Frequenzbereich vornimmt.

[0002] Daranliegende Verfahren gewinnen insbesondere für die Sprachausgabe von Kommandos und/oder für Freisprechfunktionen zunehmend an Bedeutung, wobei insbesondere die Situation in einem Fahrzeug ein wichtiger Anwendungsbereich ist.

[0003] Durch einfache Abtrennungen beim Einsatz eines Verarbeitungsschritts mit einem Zeiterster kann durch Übergang zu einer Faltung im Frequenzbereich eine deutliche Reduzierung des Verarbeitungsaufwands reduziert werden.

[0004] Für lange Impulsantworten der Übertragungsfunktion bzw. deren Nachbildung sieht eine vorlängige Weiterbildung der Erfindung eine Aufteilung des Nachbildungsfilters in mehrere Teilfilter zu zeilenseitigen Segmenten des segmentierten Referenzsignals vor, deren Koeffizienten-Aktualisierung zeitlich gestaffelt sein kann, wodurch der Signaleinverarbeitungsaufwand gering gehalten werden kann.

[0005] Als besonders vorlänglich erwies es sich, die Einstellung eines Sprachsignals auf der Basis einer Einstellung des Nachbildungsfilters, die in einer vorhergehenden Sprachpause gewonnen und gespeichert wurde, vorzunehmen.

[0006] Die Aufteilung des Nachbildungsfilters in mehrere Teilfilter und die Störberichtigung auf der Basis einer Sprachpause gewonnenen Filtereinstellung sind auch unabhängig von der Störignalkompression im Frequenzbereich eigenständig für die Störberichtigung eines Mikrophonsignals realisierbar und vorlänglich.

[0007] Die Erfindung ist nachfolgend anhand von Fig. 1 bis Fig. 8 erläutert.

## Szene Z4

fea

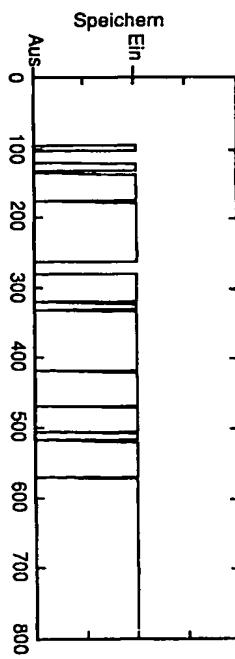
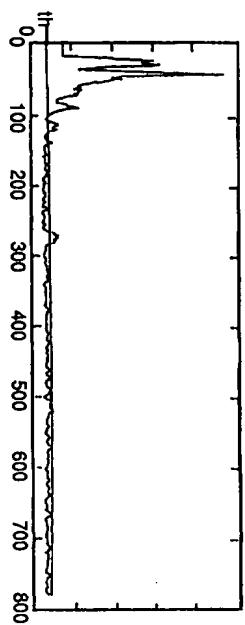


FIG.27

Fig. 1 ein Prinzip der Kompression eines Radiosignals  
 Fig. 2a ein Blockschaltbild zur Filternachbildung  
 Fig. 3 ein detailliertes Beispiel zu Fig. 2b  
 Fig. 4 eine Erweiterung auf mehrere Teilfilter  
 Fig. 5 einen Übergang zur Kompression im Frequenzbereich  
 Fig. 6 ein detailliertes Beispiel zu Fig. 5b  
 Fig. 7 ein Ausführungsbeispiel mit mehreren Teilfiltern  
 Fig. 8 ein Ausführungsbeispiel mit Speicherung

der Filtereinstellungen

Fig. 9 Signale einer synthetischen Beispielszene

Fig. 10 Impulsantwort und Übertragungsfunktion zu Fig. 9

Fig. 11 Signal einer ersten Meßszene

Fig. 12 Impulsantwort und Übertragungsfunktion zu Fig. 11

Fig. 13 das Beispiel nach Fig. 11 mit Speicherung der Filtereinstellungen

Fig. 14 eine Sprachpausendektion zu Fig. 13

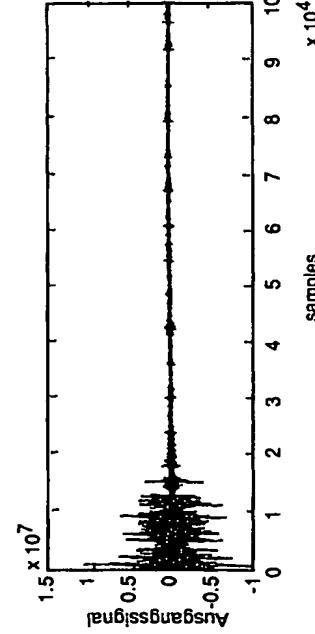
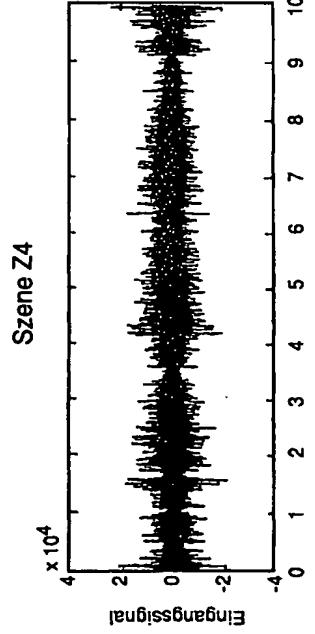


FIG.26

Sprechgeteilsystems. Unter der Annahme, daß sich die Übertragungsstrecke  $G$  danach als Transversalfilter mit einer gewichteten Summe zeitlich verzögter Echos darstellt, läßt sich eine Filternachbildung  $H$  finden, die im idealen  $H \circ G$  eine vollständige Kompensation des Radiosignals ermöglicht.

[0014] Das Lautsprechersignal  $x$  wird durch die a priori unbekannte Übertragungsfunktion  $G$  des Fahrzeuginnenraums gefiltert. Es entsteht die Störkomponente  $r$ , die sich mit dem Sprachsignal  $s$  zu dem Mikrophonesignal  $r^*$  addiert. Um die Störkomponente zu kompensieren, wird mittels der Filternachbildung  $H$  ein Schätzwert  $r^*$  aus dem Lautsprechersignal  $x$  erzeugt. Der Ausgang der Schätzung liefert den Schätzwert für das Sprachsignal:

$$s^* = s + r - r^* = s + E$$

[0015] Dem Sprachsignal  $s$  ist also am Ausgang der Schätzung noch das Fahrzeugsignal  $E = r^*$  überlagert, welches in der Praxis möglichst klein gehalten werden sollte. Das Sprachsignal kann noch Störungen in Form von z.B. Motorgetäuschen oder getarnten Geräuschen enthalten, die aber in diesem Zusammenhang nicht explizit behandelt werden.

[0016]  $H$  ist ein adaptives Filter und arbeitet nach einem in der Literatur bekannten Standardverfahren, dem LMS-Algorithmus (least mean squares). Neben dem Eingangssignal  $x$  wird noch das Fehlersignal  $E$  benötigt, um die Koeffizientenadaption im Filter  $H$  zu bewerkstelligen. Hierfür ist das Ausgangssignal  $s$  der Bestimmung der Filterkoeffizienten zugeführt.

[0017] Fig. 2a zeigt in anderer Darstellung nochmals die Anordnung von Fig. 1 als Radiosprachkomposition. Das adaptive System  $H$  kann z. B. im Zeitbereich als FIR-Filte (finite-impulse-response-Filte) realisiert werden. Bei großen Impulsantworten, wie sie in der Praxis häufig auftreten, ist hierzu allerdings ein sehr hoher Rechenaufwand notwendig. Verschiedene Varianten dieser Zeitbereichsfilterung bietet die Realisierung des LMS-Algorithmus im Frequenzbereich (F-LMS). Wegen der blockweisen Verarbeitung von Daten in den diskreten Fouriertransformationen realisierten Spektren Transformationen und der Filterrealisierung im Frequenzbereich durch Multiplikationen wird dieses Verfahren besonders rechenzeitgünstig.

[0018] Fig. 2b zeigt ein Blockschaltbild des F-LMS-Algorithmus. Die zugehörige Theorie ist an sich bekannt und daher an dieser Stelle nicht im Detail behandelt. Es bedürfen  $F$  einer spezielle Transformation FFT eines Zeitsignals in den Frequenzbereich und  $F^{-1}$  die inverse FFT. Die als Projektionen  $P_1$ ,  $P_2$  und  $P_3$  bezeichneten Verarbeitungsschritte dienen der korrekten Segmentierung der Daten durch die blockweise Verwendung mit der FFT bzw. IFFT und werden später noch genauer erläutert. Die Arbeitsweise des Filters besteht in der Multiplikation des Referenzspektrums  $X$  mit dem Filter-Koeffizientenvektor  $H$ . Das Spektrum des

Fig. 9 Signale einer synthetischen Beispielszene

Fig. 10 Impulsantwort und Übertragungsfunktion zu Fig. 9

Fig. 11 Signal einer ersten Meßszene

Fig. 12 Impulsantwort und Übertragungsfunktion zu Fig. 11

Fig. 13 das Beispiel nach Fig. 11 mit Speicherung der Filtereinstellungen

Fig. 14 eine Sprachpausendektion zu Fig. 13

Fig. 15 Impulsantworten und Übertragungsfunktionen zu Fig. 11 und Fig. 13

Fig. 16 Übergang von einem Zeitfenster zu einer Füllung im Frequenzbereich

Fig. 17 ein Rechteck-Zeitfenster mit Linienspektrum

Fig. 18 ein Hamming-Zeitfenster mit Linienspektrum

Fig. 19 Staffelung von Signalblöcken bei der Filterberechnung

Fig. 20 Signale einer zweiten Meßszene

Fig. 21 eine Sprachpausendektion zu Fig. 20

Fig. 22 Impulsantworten und Übertragungsfunktionen zu Fig. 20 und Fig. 21

Fig. 23 Signale einer dritten Meßszene

Fig. 24 eine Sprachpausendektion zu Fig. 23

Fig. 25 Impulsantworten und Übertragungsfunktionen zu Fig. 23 und Fig. 24

Fig. 26 Signale einer vierten Meßszene

Fig. 27 eine Sprachpausendektion zu Fig. 26

Fig. 28 Impulsantworten und Übertragungsfunktionen zu Fig. 26 und Fig. 27.

[0013] Fig. 1 stellt das Prinzip einer Einrichtung zur (ein kanaligen) Radiosprachkomposition dar. Das vom Lautsprecher abgestrahlte akustische Signal gelangt auf direktem Wege, aber auch über zahlreiche Reflexionen im Fahrzeuginnenraum, auf das Mikrofon des

Filtausgangs  $P_1$  wird über  $F_1^{-1}$  zurück in den Zeitbereich transformiert. Nach Anwendung der Projektion  $P_2$  auf den Realteil des so erhaltenen Kompressionssignals steht das Signal  $r^m$  zur Verfügung. Die Differenz der Signale

$$s^m = y - r^m = s + r^m = s + E$$

stellt den eigentlichen Ausgang, eine Schätzung der Spracheingabe dar.

[0019] Wesentlicher Bestandteil des adaptiven Filters ist die Koeffizientenadaption im Block  $K$ , die im Fig. 2b durch die Erneuerungsgleichung

$$H = H + \Delta H$$

beschrieben wird. Die hier mit zwei speziellen Transformationen besondere anwendende Projektion  $P_1$  berechnet aus  $H$  den für die Filterung benötigten Koeffizientenvektor  $H$ . Zur Berechnung des Korrekturvektors  $\Delta H$  wird neben dem Referenzspektrum  $X$  das Spektrum  $S^m$  des mit  $P_3$  bewerteten Ausgangssignals berechnet. Aus  $H$  den für die Filterung benötigten Koeffizientenvektor  $H$ . Zur Berechnung des Korrekturvektors  $\Delta H$  wird neben dem Referenzspektrum  $X$  das Spektrum  $S^m$  des mit  $P_3$  bewerteten Ausgangssignals berechnet.

[0020] Ein detailliertes Blockschaltbild des in Fig. 2b dargestellten FMLS-Algorithmus ist Fig. 3. Die Abstimmwerte eines Signals und die Substellen der FFT stehen in gebrauchlicher Weise als samples bezeichnet. Alle Spektraltransformationen und deren Inversen sind als 256-Punkte - FFTs, die jeweils um 128 samples überschappen, zu segmentieren. Zu beachten ist, daß sich das Ausgangssignal  $s^m$  im Zeitbereich aus 128-sample-Blocken zusammensetzt. Es entsteht aus der Differenz der zweiten Blockhälfte (also jeweils der samples 129 bis 256) von Referenzsignal und gefiltertem Kompressionsignal  $r^m$ . Anwendig ist die Projektion  $P_1$ , die 2 FFTs benötigt und den Vektor  $H$  in den Vektor  $H$  umrechnet. Hierbei wird aus dem komplexen 256-Punkte-Ergebnisvektor der Rücktransformation vom Frequenz- $\Omega$ -Bereich in den Zeitbereich die erste Hälfte (samples 1 bis 128) ausgeschnitten und die zweite Hälfte (samples 129 bis 256) zu Null gesetzt. Nach Anwendung dieses Rechteckkenners im Zeitbereich erfolgt wieder mittels FFT die Transformation in den Frequenzbereich. Einfach ist die Projektion  $P_2$ , sie besteht aus der oben schon beschriebenen Ausschnittsoperation der letzten 128 samples, wodurch aus überlappenden 256-sample-Blocken wieder nicht überlappende 128-sample-Blocke entstehen. Eindeutig sehr einfach ist schließlich auch die Projektion  $P_3$ , welche ungestört aus nicht überlappenden 128-sample-Blocken des Ausgangssignals durch Vorausstellen von 128 Nullwerten wiederum überlappende Sample-Blocke bereitstellt. Die Adaption der Filterkoeffizienten  $H_{L+1}$  für einen Zyklus  $L+1$  besteht aus der Addition eines Erneuerungsspektros  $\Delta H$  zum alten Koeffizientenvektor  $H_L$ . Diese Erneuerung erreicht sich aus dem Produkt zwischen dem Spektrum  $S^m$  des Ausgangssignals und dem konjugiert komplexen Spektrum  $X_L^*$  des Referenz-

signals, gewichtet mit einer speziellen Leistungsmaß  $2H_L \cdot \Delta H = 2H_L \cdot X_L^* \cdot S^m$ . Zum Zweck dieser Leistungsmaßberechnung ist der mit einer Konstanten  $\alpha$  multiplizierte Kehrwert des geballten Referenzleistungsspektrums  $S_{xx,L}$  zu berechnen

$$S_{xx,L} = \beta \cdot [X_L]^2 + (1-\beta) \cdot S_{xx,L-1}$$

10

15

20

25

30

35

40

45

50

55

60

65

70

75

80

85

90

95

100

105

110

115

120

125

130

135

140

145

150

155

160

165

170

175

180

185

190

195

200

205

210

215

220

225

230

235

240

245

250

255

260

265

270

275

280

285

290

295

300

305

310

315

320

325

330

335

340

345

350

355

360

365

370

375

380

385

390

395

400

405

410

415

420

425

430

435

440

445

450

455

460

465

470

475

480

485

490

495

500

505

510

515

520

525

530

535

535

540

545

550

555

560

565

570

575

580

585

590

595

600

605

610

615

620

625

630

635

640

645

650

655

660

665

670

675

680

685

690

695

700

705

710

715

720

725

730

735

740

745

750

755

760

765

770

775

780

785

790

795

800

805

810

815

815

820

825

830

835

840

845

850

855

855

860

865

870

875

880

885

890

895

900

905

910

915

920

925

930

935

940

945

950

955

960

965

970

975

980

985

990

995

1000

1005

1010

1015

1020

1025

1030

1035

1040

1045

1050

1055

1060

1065

1070

1075

1080

1085

1090

1095

1095

1100

1105

1110

1115

1120

1125

1130

1135

1140

1145

1150

1155

1160

1165

1170

1175

1180

1185

1190

1195

1200

1205

1210

1215

1215

1220

1225

1230

1235

1240

1245

1250

1255

1260

1265

1270

1275

1280

1285

1290

1295

1295

1300

1305

1310

1315

1320

1325

1330

1335

1340

1345

1350

1355

1360

1365

1370

1375

1380

1385

1390

1395

1400

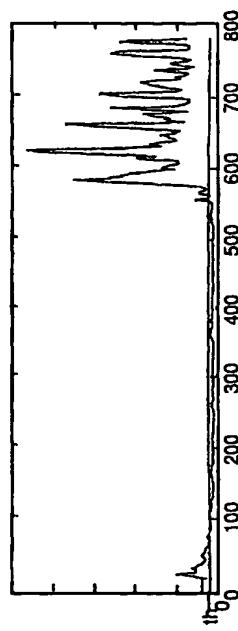
1405

1410

1415

$y$  ist nur eine einzige 256-Punkte-FFT mit geringem Zusatzaufwand für eine spektrale Separation notwendig. Man erzielt eine Einspannung von 1 FFT.

Die hier mit P4 gekennzeichnete und neu definierte Projektion ist auf das verwendete Zeitfenster formal identisch mit der Projektion P1. Wie später gezeigt wird, läßt sich P4 durch eine relativ einfache Faltungsoperation im Frequenzbereich ersetzen, ohne daß eine merkliche Einbuße an Qualität in Kauf genommen werden muß. Man erzielt eine Einspannung von 2 FFTs.



24a

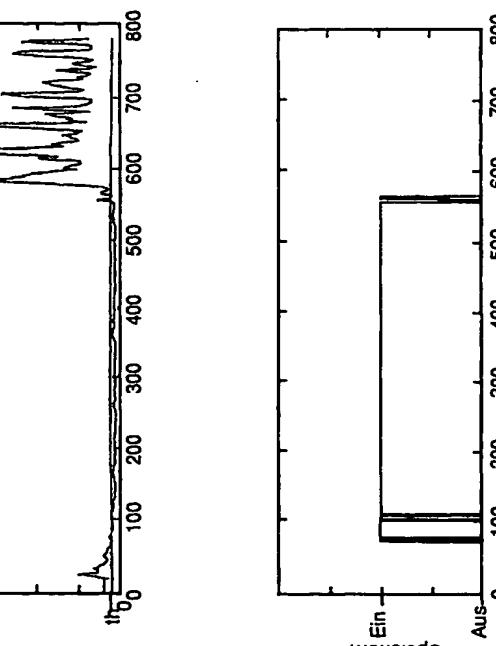


FIG.24

gefilterten Rauschen zu erkennen. Das in den Zeitbereich zurücktransformierte Ausgangssignal des Schätzers befreit nach einem ca. 1 Sekunde (12000 samples) dauernden Einschwingvorgang sehr wirkungsvoll die Sprachengebiete vom Rauschen und liefert ein unverzerrtes, jedoch leicht verhältnisses Sprachsignal  $S^*$  (Fig. 9 unten). Die verwendeten beiden Parameter waren  $\alpha=0.05$  und  $\beta=0.5$ . Werte, die sich auch bei dem später vorgestellten Beispiel gut bewährt haben.

[0027] Aus den jeweils 128 samples langen Teilkoefizientenvektoren  $H_1$ ,  $H_2$ ,  $H_3$  der 3 Teilfilter nach Bild 7 läßt sich nun zu jedem Zeitpunkt die hieraus resultierende 3128-sample-Impulsantwort berechnen. So zeigt Fig. 10 oben die 384-sample-Impulsantwort, wie sie sich ganz am Ende der Szene also nachdem die Ziffer '0' gesprochen wurde - erzielt. Sie ist ein sehr glattes Abbild der ursprünglichen Impulsantwort, die zur Faltung mit weißem Gaußrauschen und damit zur synthetischen Generierung des Signales mikro benutzt wurde. Die zugängliche Betriebsübertragungsfunktion (Fig. 10 unten) im Bereich zwischen den Frequenzen 0 und  $f_s/2 = 6$  kHz stellt einen mit zahlreichen schnabelförmigen Resonanzüberhöchungen behafteten Trichterfrequenzgang dar.

[0028] Weißes Rauschen als Referenzeingangssignal und genügend "lauriges" Rauschen als Mikrophoneingangssignal stellen im Sinne der Aufgabenstellung eine Nachbildung dieses Filters zu finden, den einfacheren Fall dar. Da das Referenzsignal per Definition alle Frequenzanteile enthält, gelingt die Filteradaption hier am schnellsten. Die zusätzliche additive Spracheneingabe im Mikrophoneingangssignal - also das eigentliche Nutzsignal des Sprachreinigungsabgesystems - stellt für den (FLMS)-Algorithmus eine Störung dar, welche die korrekte Adaption der Filterkoefizienten behindert. Anders ausgedrückt: nur in Sprachpausen ist das System dazu in der Lage die Raumakustik des Fahrzeuginnenraums (Strecke: Radiolautsprecher zum Mikrofon) richtig nachzubilden und dadurch eine Kompensation der Radiowiedergabe zu bewirken. Im dem demonstrierten Beispiel (siehe Fig. 9) gelingt dies sehr gut, da die Mikrofon-Engstigung im wesentlichen aus Rauschen und nur zu einem sehr geringen Teil aus Sprachengebiete besteht.

[0029] Aus echten Messungen im Fahrzeug hängen entstammt das an den Radiolautsprecherketten abgeleitete Referenzsignal radio und das vom Mikrofon des Sprachreinigungsabgesystems aufgezeichnete Signal mikro der Szene 21. Dieses Mikrophonesignal ist in Fig. 11 oben dargestellt, besteht aus 100000 samples und besitzt demnach bei einer Sampling-Frequenz von 12 kHz eine zeitliche Dauer von ca. 8.3 Sekunden. Es handelt sich um flüssig und relativ schnelle gesprochene Sprache eines 'hinter rechts im Auto sitzenden Fahrzeuginsassen, während gleichzeitig Musik mit normaler Lautstärke aus dem Autoradio-Lautsprecher ertönt. Nach Anwendung der Entstörmaßnahmen gemäß Fig. 7

und Umrechnung in den Zeitbereich ergibt sich das in Fig. 11 unten dargestellte Ausgangssignal. Der Moment eines vor allem in den kurzen Sprachpausen eingeschlossene Muskunterscheidung, Auftakt und von

Nachteil ist jedoch, daß die erwartete Radiosignalunterdrückung in starkem Maße davon abhängt, ob gerade gesprochen wird oder nicht. Die wieder am Sprechende ermittelte 384-Sample-Impulsantwort mit zugehöriger Übertragungsfunktion ist aus Fig. 12 ersichtlich. Eine korrekte Impulsantwort ist an den typischen Nullsamples (Tritzen) am Anfang zu erkennen, welche von der Laufzeit des Direktschalls vom Radiosprecher zum Mikrofon herführen. Aus dem hier laufendenen starken Störungen am Anfang sowie am Ende der Impulsantwort läßt sich dermaßen die Schlußfolgerung, daß die Filteradaption an dieser Stelle wegen vorhandener Sprachsingabe außerst unzureichend ist.

[0330] Die im folgenden aufgrund von Fig. 8 beschriebene Ausführungsform beruht auf folgender Grundidee: ein gegebenes Merkmal dient zusammen mit einem Schwellenwert als Indikator für eine Sprachsingabe. Unterscheidet das Merkmal die Sprachzeit, so ist dies ein Anzeichen für fehlende Sprachsingabe. In diesem Fall kann - wie oben schon festgestellt - eine weitgehend ungünstige Filteradaption erfolgen. Bei Sprachsingabe wird nun aus denjenigen Filterkoeffizientenwerten zurückgegriffen, der unmittelbar vor der Schwellentrennscheide - d. h. am Ende der vorangegangenen Sprachpause - abgespeichert wurde. Diese gespeicherten Koeffizienten H10, H20, H30 liefern im Regelfall eine deutlich bessere Radiosignal-Kompensation als die unter dem ständigen Einfluß der Sprachsingabe sich ständig ändernden aktuellen Koeffizienten H1, H2.

[0331] Fig. 8 stellt eine Ausführung mit einer weiter verbesserten FLMs-Verarbeitung mit 3 Tritten dar. Neben den schon in Fig. 7 veränderten aktuellen Filterkoeffizientenwerten H1, H2, H3, welche zur Bildung des fortlaufend adaptierten Ausgangssignals y-R benötigt wurden, existiert nun ein zusätzlicher Ausgangssignal (y-Ro), das unter Verwendung gespeicherter Koeffizienten H10, H20, H30 gebildet ist. Die aktuellen Koeffizientenwerte H1, H2, H3 stellen nur bei fehlender Sprachsingabe im eingeschwungenen Zustand ein brauchbares Kompressionsmittel im Frequenzbereich dar, liefern hingegen bei Sprachsingabe ungünstige Filtereigenschaften, weil der Adaktionsprozeß in der Regelschleife ständig gestört wird. Bei fehlender Sprachsingabe d. h. hoher Filterqualität sind die drei Schalter geschlossen und es werden die aktuellen Koeffizientenwerte in die 3 Schalter M1, M2, M3 geschrieben: H10=H1, H20=H2, H30=H3. Die Ausgänge (y-Ro) und (y-Ra) sind identisch. Einsatzende der Sprachsingaben beruhten der Offnen der 3 Schalter, wodurch die zuletzt in den Speichern M1, M2, M3 befindlichen Koeffizienten H10, H20, H30 nicht mehr überschrieben werden und unverändert bleiben. Dieser

Zustand, in welchem sich die Ausgänge (y-Ro) und (y-Ra) unterscheiden, wird solange beibehalten, bis wieder geschlossen werden.

[0332] Als Sprachpausenmerkmal fea hat sich die begrenzte Summe aller Abschlußwerte der Koeffizientenkonkurrenzarten  $\Delta H1^1, \Delta H2^1, \Delta H3^1$  bewährt (Fig. 8a). Diese Größe ist gleich Null bzw. weist kleine Zahlenwerte auf, wenn es keinen oder nur einen geringen Bedarf gibt, die Koeffizienten abzuladen. In Sprachpausen ist dies der Fall, der Regelkreis ist praktisch einschlafen. Als Sprachpausenmerkmal fea hat sich die begrenzte Summe aller Abschlußwerte der Koeffizientenkonkurrenzarten  $\Delta H1^1, \Delta H2^1, \Delta H3^1$  bewährt (Fig. 8a).

Diese Größe ist gleich Null bzw. weist kleine Zahlenwerte auf, wenn es keinen oder nur einen geringen Bedarf gibt, die Koeffizienten abzuladen. In Sprachpausen ist dies der Fall, der Regelkreis ist praktisch einschlafen. Als Sprachpausenmerkmal fea hat sich die begrenzte Summe aller Abschlußwerte der Koeffizientenkonkurrenzarten  $\Delta H1^1, \Delta H2^1, \Delta H3^1$  bewährt (Fig. 8a).

Ausgangssignal

Scene Z3

Eingangssignal

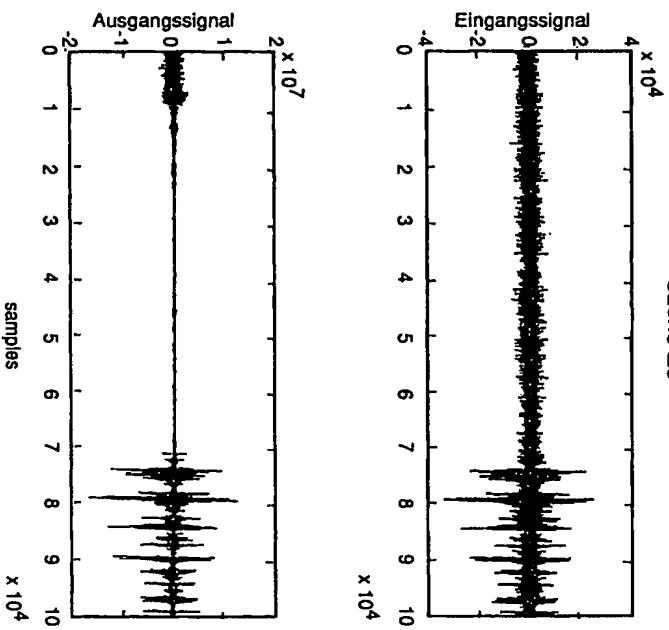


FIG.23

[0332] Als Sprachpausenmerkmal fea hat sich die begrenzte Summe aller Abschlußwerte der Koeffizientenkonkurrenzarten  $\Delta H1^1, \Delta H2^1, \Delta H3^1$  bewährt (Fig. 8a). Diese Größe ist gleich Null bzw. weist kleine Zahlenwerte auf, wenn es keinen oder nur einen geringen Bedarf gibt, die Koeffizienten abzuladen. In Sprachpausen ist dies der Fall, der Regelkreis ist praktisch einschlafen. Als Sprachpausenmerkmal fea hat sich die begrenzte Summe aller Abschlußwerte der Koeffizientenkonkurrenzarten  $\Delta H1^1, \Delta H2^1, \Delta H3^1$  bewährt (Fig. 8a).

[0333] Die Wirkungsweise des verbesserten FLMs-Algorithmus nach Fig. 8 demonstriert Fig. 13. Oben ist das aufgezeichnete Signal u (der Szene Z1 (vgl. Fig. 11 oben)) dargestellt, unten das gewonnene Ausgangssignal. Schon der visuelle Vergleich der Ausgangssignale zeigt Fig. 13 und Fig. 11 zeigt die verbesserte Herausarbeitung der Sprachpassagen. Der vergleichende Test bestätigt dies: auch während der Sprachsingabe ist die Musikunterdrückung deutlich besser. Den Verlauf des Sprachpausenmerkmals und der konstanten Schwellen über der Zeit (hier in FFT-Blöcken skaliert) zeigt Fig. 14 oben. In den durch die Schwellenunterstützung definierten Sprachpausen (Fig. 14 unten) findet laufend die Übernahme der Koeffizienten in die Sprachzeit über der Zeit (hier in FFT-Blöcken skaliert) statt, um dort während der Sprachsingabe als gespeicherte Koeffizienten zur Verfügung zu stehen. Die schon in Fig. 12 am Szeneende gemessene 384-Sample-Impulsantwort mit Lugehöriger Betragssubtraktionsfunktion ist in Fig. 15 als aktuelle Impulsantwort (a) bzw. aktuelle Übertragungsfunktion (b) dargestellt. Im Gegensatz zu dieser infolge Sprach-Singabe stark gestörten Schätzung aus den aktuellen Koeffizienten H1, H2, H3 ist aus den gespeicherten Koeffizienten H10, H20, H30 eine Impulsantwort (c) und eine Übertragungsfunktion (d) hoher Qualität berechenbar. Die Impulsantwort aus den gespeicherten Koeffizienten weist die typischen Nullsamples am Anfang auf, welche durch die Laufzeit des Direktschalls vom Radiosprecher zum Spracheingangsmikrofon verursacht werden. Aus der im Beispielstall abzuzeigenden Totzeit von ca. 40 samples läßt sich die Entfernung zwischen Lautsprecher und Mikrofon bestimmen.

[0334] Wie vorstehend schon angeführt läßt sich die authentische Projektion P4 (IFFT, Fenster rechts im Zeitbereich, FFT) ohne merkliche Einbuße an Qualität

durch eine relativ einfache Faltung im Frequenzbereich ersetzten, wodurch 2 FFT's eingespart werden. Dazu betrachte man Fig. 16. In einem ersten Schritt wird das "rechteckige" 128-sample-Rechteckfenster im Zeitbereich (Fig. 16a) bei der idealen Projektion ersetzt durch ein 128-sample-Hammingfenster (Fig. 16b). Gegenüber dem Rechteckfenster besitzt dieses den Vorteil eines bedeutend schmaleren Spektrums. Wie Fig. 17 zeigt, besteht beim Rechteckfenster der Projektions-Spektrum aus einer einzigen Linie (Gleichanteil), während das zur Mitte antisymmetrische Imaginärteil-Spektrum aus vielen nach außen hin langsam absinkenden Linien mit alternierenden Nullstellen besteht. Im Gegensatz dazu beschreibt sich das komplexe Spektrum des Hammingfensters (Fig. 16b) auf insgesamt 7 Linien, von denen im symmetrischen Realteil nur 3 und im antisymmetrischen Imaginärteil nur 4 Werte von Null verschieden sind. Sämtliche weiter außen liegenden Anteile sind vernachlässigbar gering. Diese spezielle Eigenschaft des Hammingfensters ermöglicht es vor teilhaftweise die Multiplikation im Zeitbereich (Fig. 16c) zu ersetzen durch eine Faltung mit dem zugehörigen 7-sample-Spektrum im Frequenzbereich und damit eine IFFT und eine FFT einzusparen (Fig. 16c).

[0005] Prinzipiell lässt sich natürlich auch die Projektion P1 (IFFT - linkezeitiges Rechteckfenster - FFT) ersetzen durch eine entsprechende Faltungoperation im Frequenzbereich mit dem kompliziert komplexen 7-Linien-Spektrum. Experimente haben jedoch gezeigt, dass Einsparungen an dieser Stelle erwartet werden müssen. Eine deutliche Verschlechterung des Einschwingverhaltens. Aufwandsgünstige Lösungen lassen sich trotzdem dadurch erzielen, dass in dem LMS-Algorithmus nach Fig. 8 die 3 Projektionen P1 nicht gleichzeitig in einem 256-sample-Inputdatenblock abgearbeitet werden müssen. Die mit 128-samples überlappenden Inputdatenblöcke der Länge 256 sind mit einer willkürlich bei "1" beginnenden Nummerierung in Fig. 19a skizziert. So ist es z.B. möglich bei modulo-3-Zähweise die Inputdatenblöcke die 3 Teilfilterprojektionen nicht parallel (Fig. 19b) sondern sequentiell in aufeinanderfolgenden Blöcken Fig. 19 zu berechnen. Dadurch sind bei idealer Projektion P1 pro Datenblock nicht 6 sondern nur noch 2 FFT's notwendig. Es hat sich gezeigt, dass die Kompensation des Radarsignals auch noch ausreichend funktioniert, wenn die Abstände zwischen den zu berechnenden Teilfilter-Projektionen noch größer gewählt werden. Zählt man die Blöcke z.B. modulo 6, so ist lediglich in jedem zweiten Block eine Projektion zu berechnen (Fig. 19d). Selbst eine Reduzierung auf einen Abstand von vier Blöcken zwischen zwei aufeinanderfolgenden P1-Berechnungen mittels modulo-12-Zählung führt noch zu brauchbaren Ergebnissen (Fig. 19e).

[0006] Die Leistungsfähigkeit des LMS-Algorithmus mit 3 Teilfiltern gemäß Blockschaltung Fig. 8 und einer sequentieller Berechnung der idealen Projektion P1 im Zeitraffer nach Fig. 19e sowie der Projektion P2 mittels

Faltung im Frequenzbereich (Fig. 16c) mit einem komplexen 7-Linien-Spektrum (Fig. 18) sei anhand von 3 Meßszenen demonstriert.

[0037] Die erste dieser Szenen 22 beinhaltet Spracheingabe von Ziffern, wobei der Radiotalkansprecher anhendend weißes Rauschen mit variabler Intensität, hoher Lautstärke abstrahlt. Das zugehörige 100000-sample-Mikrophonesignal ist in Fig. 20 oben, das entartete Ausgangssignal in Fig. 20 unten dargestellt. Eine deutliche Rauschentfernung des Outputsignals gegenüber dem Mikrophonesignal stellt man durch Abhörengleich fest. Der zeitliche Verlauf des Sprachpausenmerkmals ist zusammen mit den konstanten Schwellen in Fig. 21 oben abgebildet und die hieraus abgeleiteten Sprachpausen bzw. die zugehörigen Schaltstellungen in Fig. 21 unten. Schließlich zeigt Fig. 22 in zu Fig. 15 analoher Weise die am Szenenende gefundene Impulsantwort (a) und Übertragungsfunktion (b) auf der Basis der aktuellen Konfiguration und die entsprechenden Größen (c), (d) auf der Basis der Sprachpauseneinstellung. Es ist deutlich erkennbar, dass die am Szenenende gefundene aktuelle Impulsantwort ein insoweit Sprachengabe gestortes Ergebnis darstellt, während die aus der letzten Sprachpause stammende Impulsantwort aus den gespeicherten Koeffizientensätzen eine hohe Qualität aufweist.

[0038] Die ersten 100000 samples einer Meßszene 23 mit POP-Musik im Radio und flüssig bis schnell gesprochener Sprache der rechts hinten sitzenden Person sind in Form des Mikrophonesignals y in Fig. 23 oben aufgezeichnet. Nach ca. 10000 samples (0.83 s) wird das Radiosignal brauchbar unterdrückt (Fig. 23 unten). Auch bei der im letzten Drittel dieser Szene eingesetzten Spracheingabe bleibt die POP-Musikunterdrückung wirksam erhalten, wodurch die Sprachverständlichkeit hier gegenüber dem Mikrophonesignal merklich verbessert wird. Nach einer langen Sprachpause kommt es wegen der anschließenden pausenfreien Spracheingabe nicht mehr zu einer Schwellenunterschreitung (Fig. 24). Aus diesem Grunde ist in Fig. 25 unten am Ende der Szene festgehaltene Impulsantwort auf der Basis der gespeicherten Koeffizienten zeitlich relativ veraltet, weil sie bereits 2.3 s vorher aktuell war (215 Blöcke \* 10.7 ms). Wie ca. der weist die aktuelle Impulsantwort (Fig. 25 oben) die Spracheingabe handhabende Störungen auf. Wie ein Vergleich mit der ähnlichen Szene 21 nach Figuren 11 bis 15 zeigt, ist trotz des stark veränderten Rechenauwandes die Qualität der Störbefreiung unverändert hoch.

[0039] Die letzte Szene 24 nach Fig. 26 wurde ohne Spracheingabe erstellt und soll abschließend nochmals die Musikunterdrückungseigenschaften des beschriebenen FMLS-Algorithmus demonstrieren. Nach ca. 16000 samples bzw. 1.5 s wird - wie aus Fig. 26 unten ersichtlich - die Musik wieder unterdrückt. Diese Eigenschaft wird bis zum Szenenende mit unveränderter Qualität beibehalten. Fig. 27 zeigt auf, dass das

## Szene 22

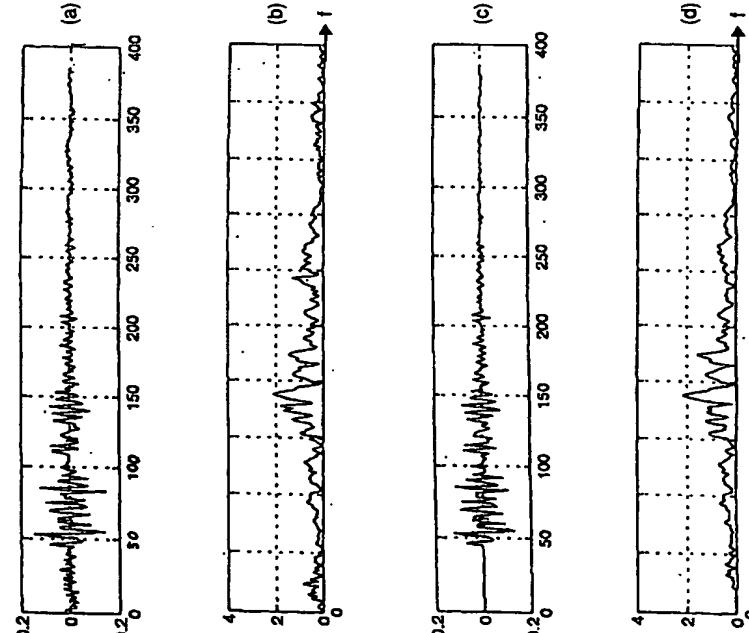


FIG.22

Sprachpausen-Größe  $t_{pa}$  überwiegend unter der Schwellle  $t_h$  bleibt. Die Zeiten, in welchen nur die gespeicherten Koeffizienten zurückgegriffen wird, sind demnach nur sehr kurz. Impulsantwort und Übertragungsfunktion aus aktuellen Koeffizienten sind daher im wesentlichen mit den entsprechenden Verläufen aus Sprachpausen-Koeffizienten identisch.

#### Patentansprüche

1. Verfahren zur Störfeierung eines Mikrohonsignals von Anteilen eines Quellsignals, daß ein Referenzsignal ( $x$ ) vorliegt und nach Durchlaufen einer Übertragungstrecke mit a priori unbekannter Übertragungsfunktion ( $G$ ) sich im Mikrohonsignal die Störung ( $l$ ) einem Sprachsignal ( $s$ ) überlagert, durch adaptive Nachbildung des Störsignals und Komensation des tatsächlichen und des nachgebildeten Störsignals in einem Ausgangsgral, wobei das Mikrohonsignal gleichfalls in dem Frequenzbereich transformiert, die Kompositionen und das im Frequenzbereich vorliegende Ausgangssignal zur Adaptation der Nachbildung mit dem im Frequenzbereich vorliegenden Referenzsignal verknüpft wird.
2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß das Ausgangssignalspektrum in den Zeitbereich transformiert, das Zeitsignal durch Voranstellen von Nullwerten auf doppelte Länge gebracht, in den Frequenzbereich rücktransformiert und der Nachbildung der Übertragungsfunktion zugrundegelegt wird.
3. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß das Ausgangssignalspektrum mit dem Spektrum eines Hamming-Zeitfensters geteilt wird und der Nachbildung der Übertragungsfunktion zugrundegelegt wird.
4. Verfahren nach einem der Ansprüche 1 bis 3, dadurch gekennzeichnet, daß zur Nachbildung des Störparameters eine adaptive Filterfunktion eines Nachbildungsfilters auf das Referenzsignal angewandt wird.
5. Verfahren nach Anspruch 4, dadurch gekennzeichnet, daß die Filterfunktion durch einen Koeffizientenvektor vorgegeben wird, dessen Koeffizienten adaptiv eingestellt werden.
6. Verfahren nach Anspruch 4 oder 5, dadurch gekennzeichnet, daß das Auftreten eines Sprachsignals im Mikrohonsignal detektiert wird und bei Auftreten eines Sprachsignals die vor Auftreten des Sprachsignals eingesetzte Filterfunktion zur Bildung des Ausgangssignals beibehalten wird.
7. Verfahren nach Anspruch 6, dadurch gekennzeichnet, daß auch bei Detektion eines Sprachsignals die adaptive Nachbildung einer aktuellen Filterfunktion zusätzlich zur Bildung des Ausgangssignals fortgeführt wird.
8. Verfahren nach Anspruch 7, dadurch gekennzeichnet, daß das Auftreten eines Sprachsignals aus einer Veränderung der aktuellen Filterfunktion detektiert wird.
9. Verfahren nach Anspruch 8, dadurch gekennzeichnet, daß die Veränderung der aktuellen Filterfunktion für die Detektion des Auftretens eines Sprachsignals zeitlich geglättet wird.
10. Verfahren nach einem der Ansprüche 4 bis 9, dadurch gekennzeichnet, daß die Filterfunktion in mehrere Teillfilterfunktionen zu aufeinanderfolgenden Abschnitten einer Gesamt-Impulsantwort aller Teilfilter aufgespalten ist und auf Referenzsignalspektrum aufgeteilten und auf Referenz-Zeitsignalen segmentierten Referenz-Zeitsignalen angewandt wird.
11. Verfahren nach Anspruch 10, dadurch gekennzeichnet, daß die Adaptation der Filterfunktion für die Teillfilter parallel durchgeführt wird.
12. Verfahren nach Anspruch 10, dadurch gekennzeichnet, daß die Adaptation der Filterfunktion für die einzelnen Teillfilter zeitsequentiell durchgeführt wird.

#### Szene 22

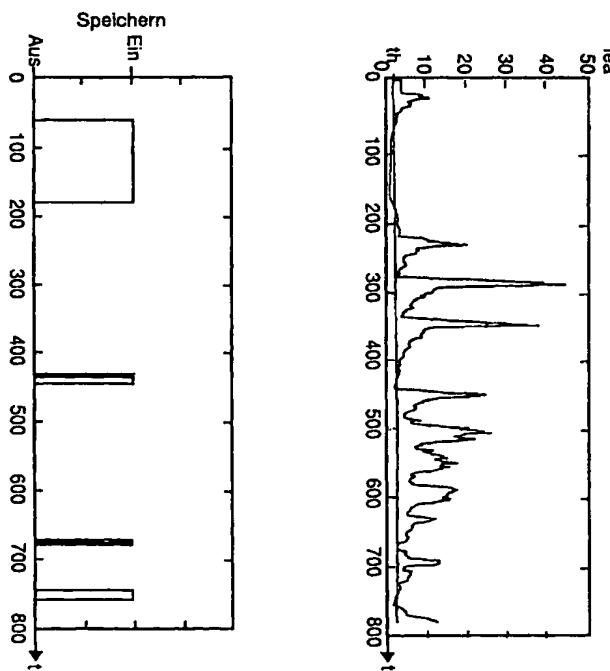


FIG.21

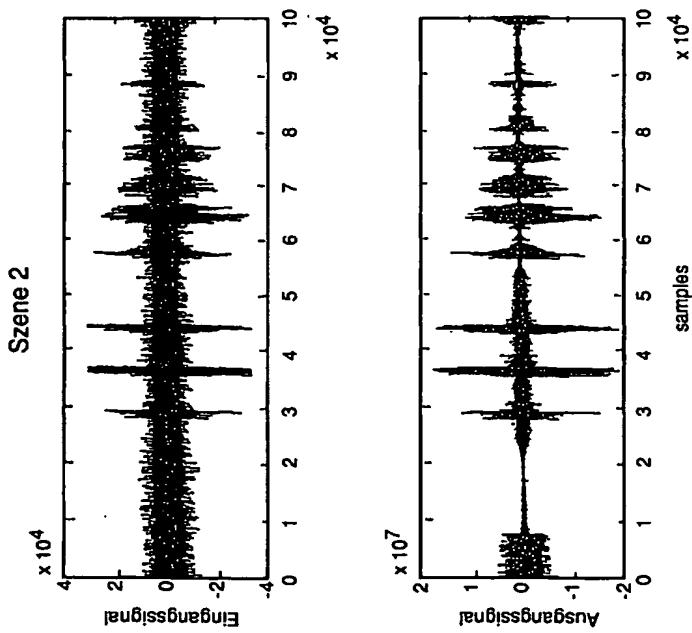


FIG.20

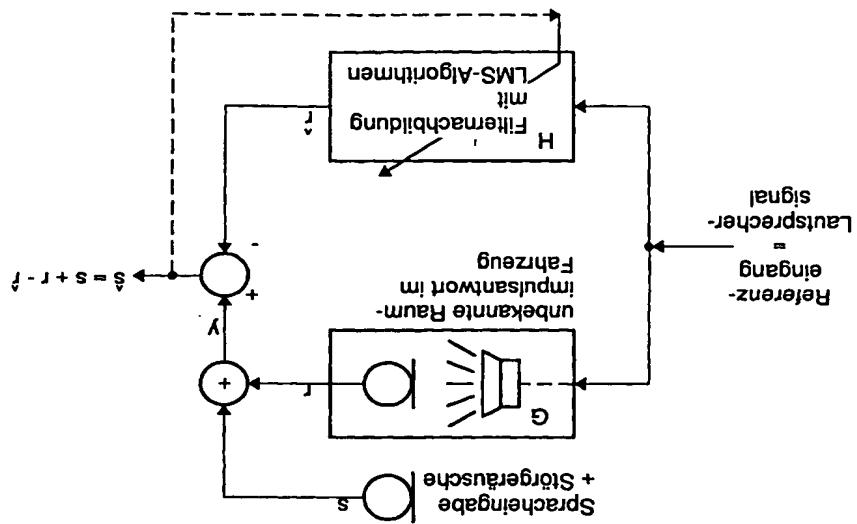


FIG.1

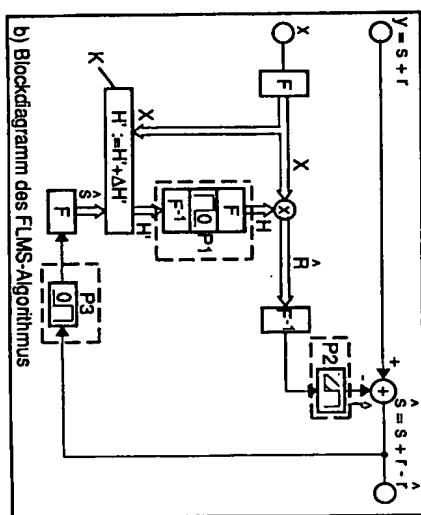
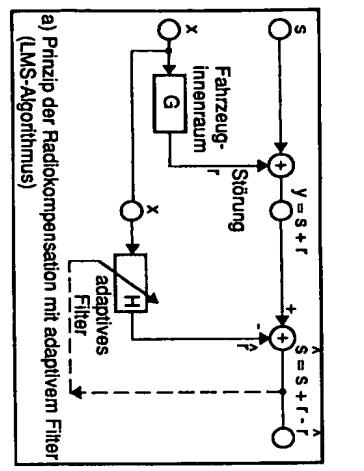
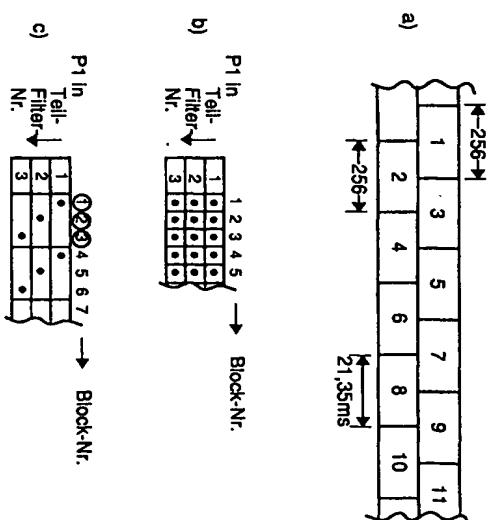


FIG.2

FIG.19



- Nummerierung der überlappenden FFT-Blöcke
- Projektion  $P_1$  parallel in 3 Teillfiltern
- Projektion  $P_1$  sequentiell in 3 Teillfiltern (Blockmodul 3)
- Projektion  $P_1$  sequentiell in 3 Teillfiltern (Blockmodul 6)
- Projektion  $P_1$  sequentiell in 3 Teillfiltern (Blockmodul 12)

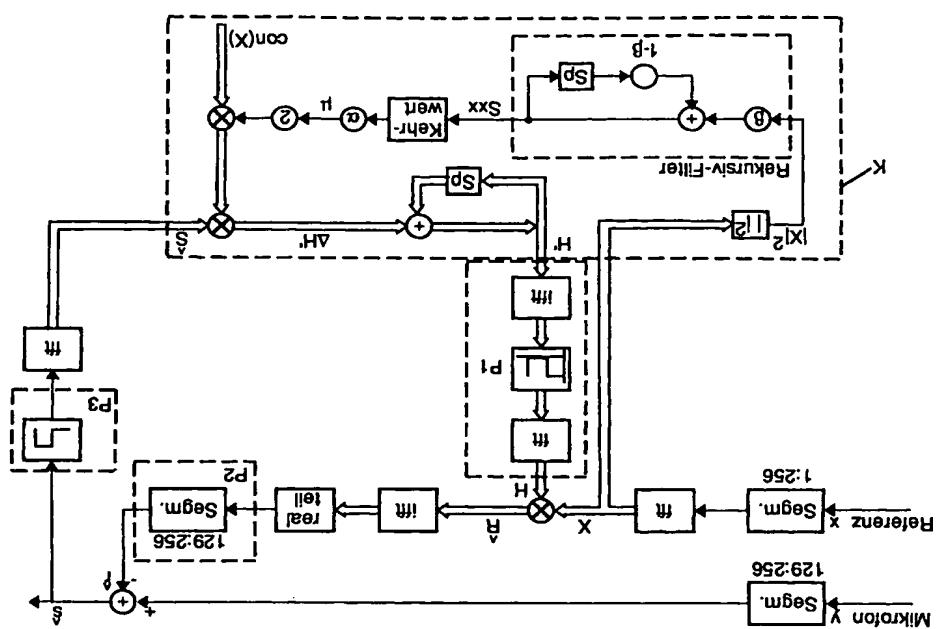


FIG.3

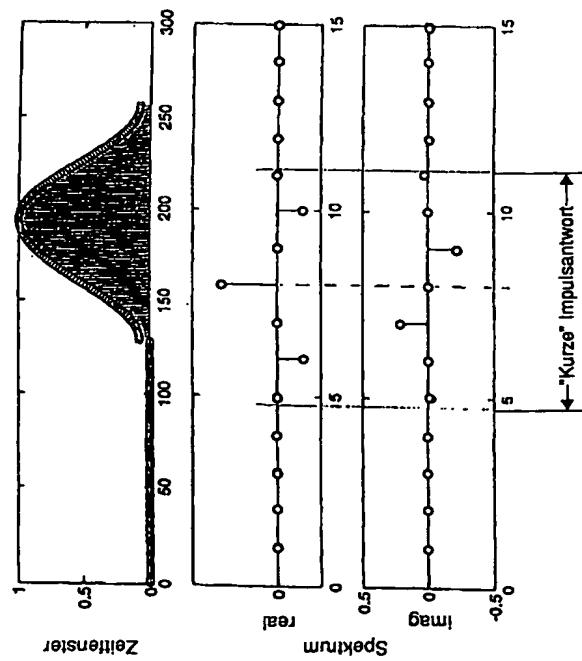


FIG. 18

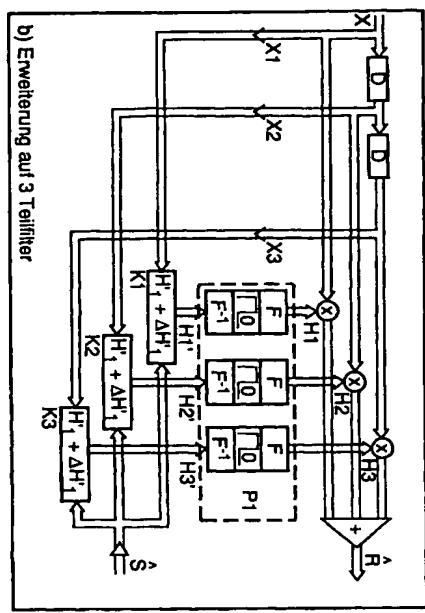
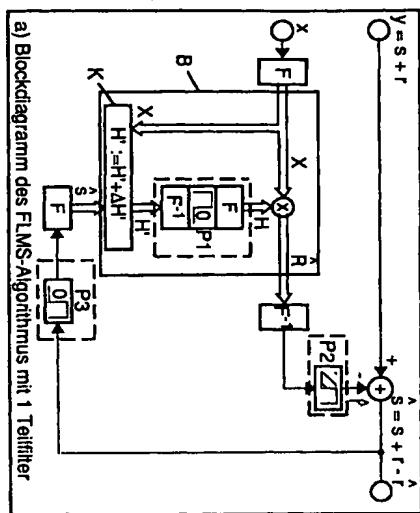


FIG.4

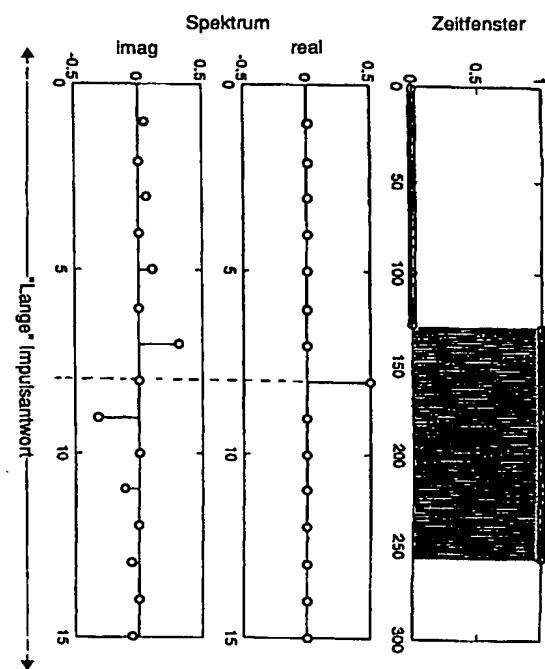


FIG.17

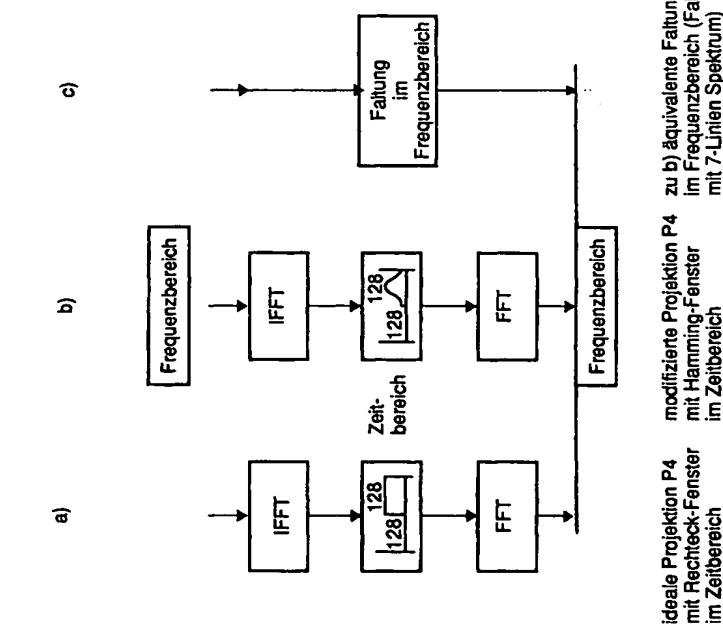


FIG.16

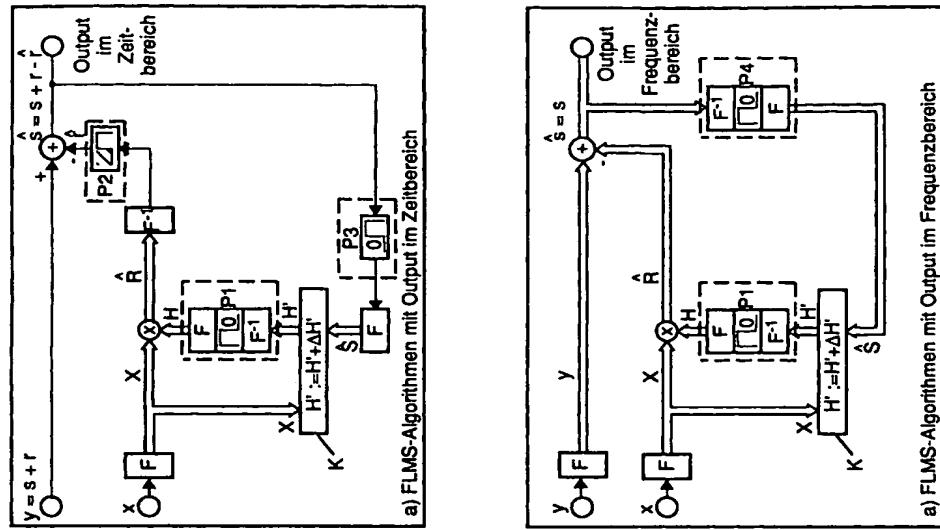


FIG.5

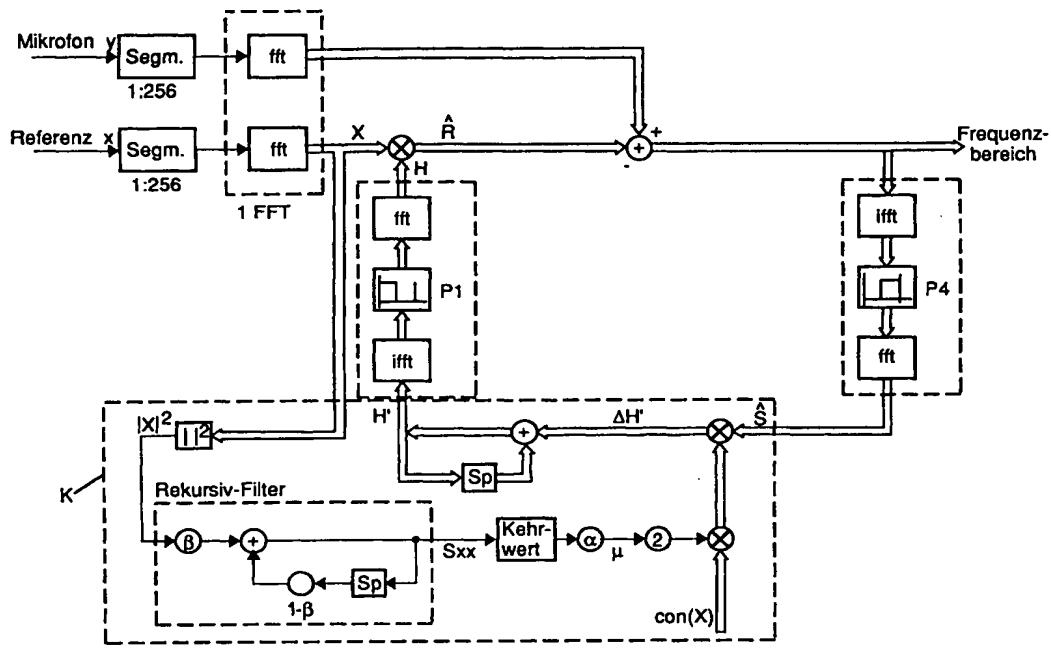


FIG.6

14

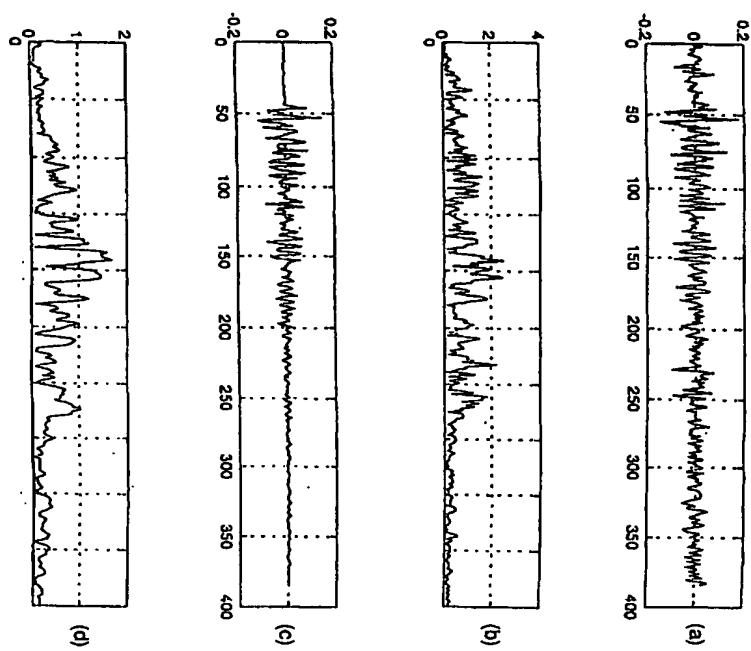


FIG.15

23

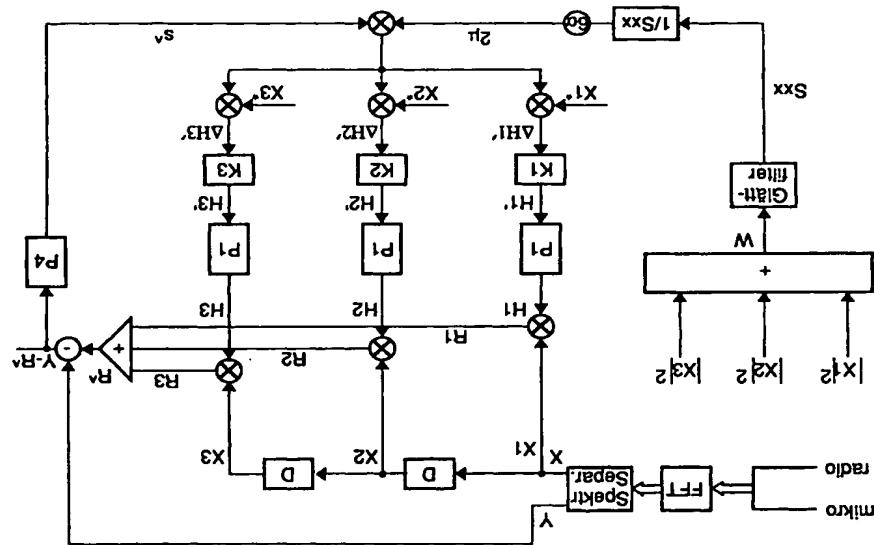


FIG. 7

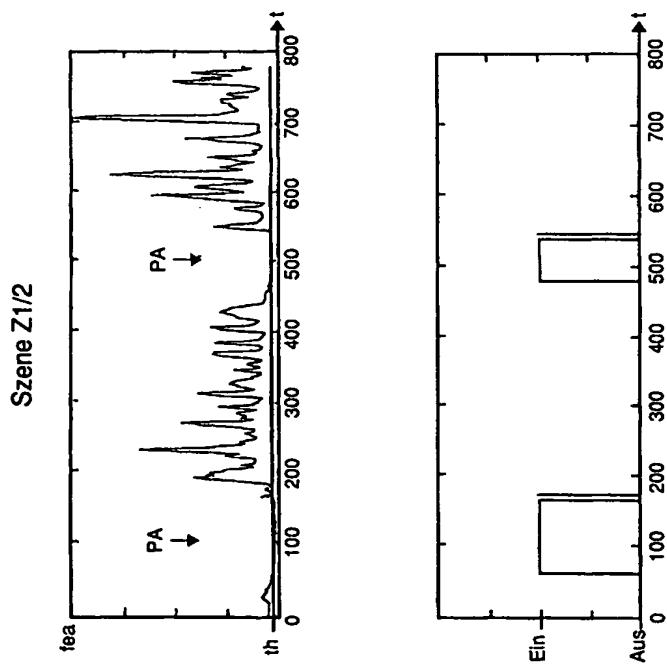


FIG. 14

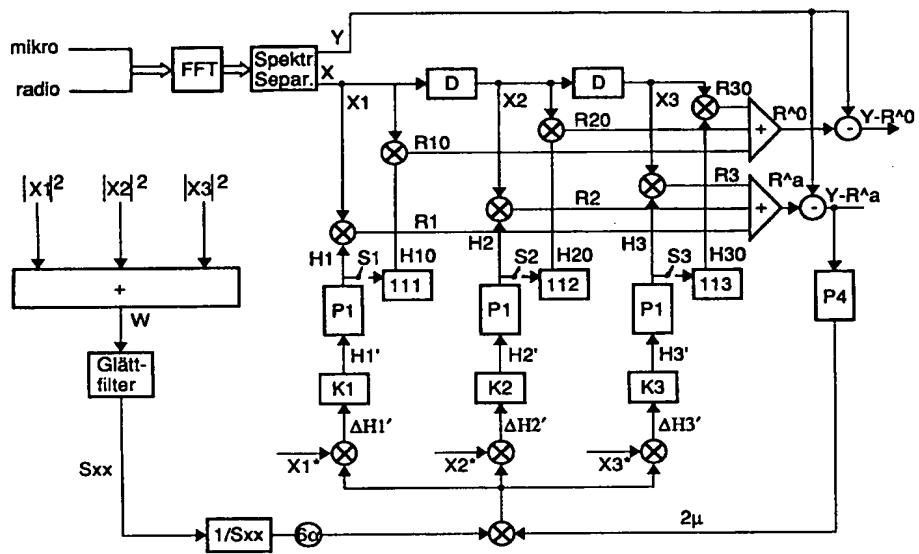


FIG.8

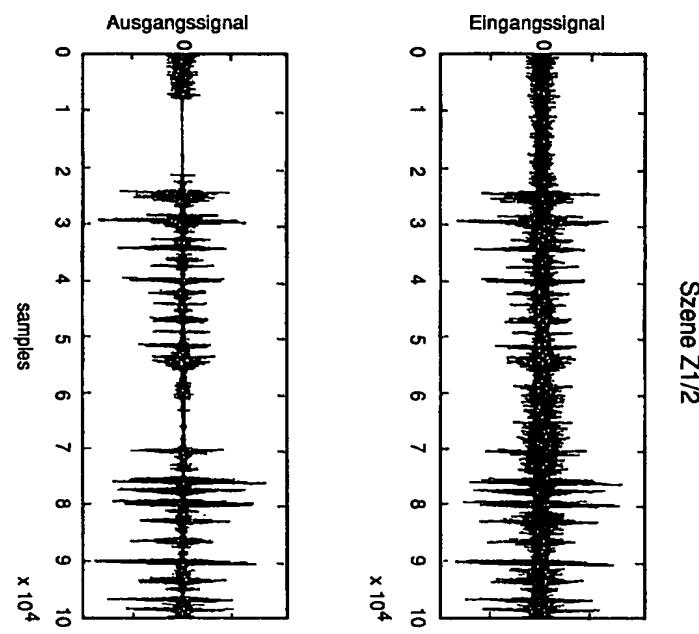


FIG.13

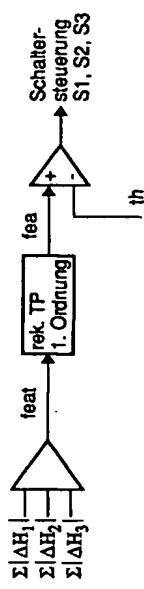


FIG.8a

Szene Z0

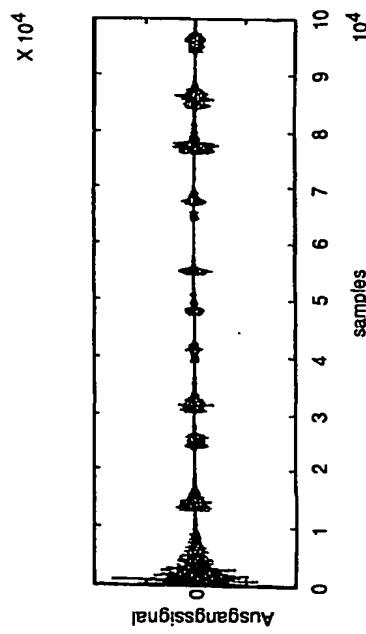
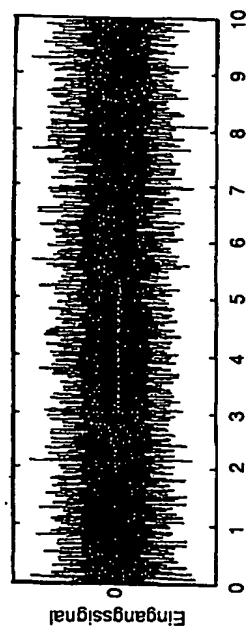


FIG.9

Szene Z1/1

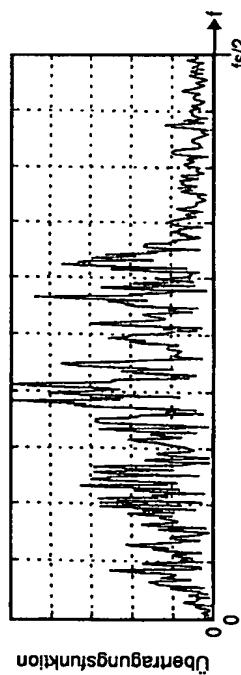
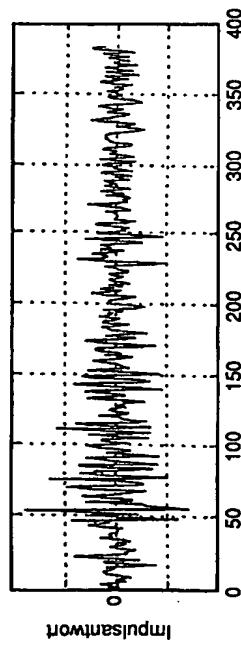


FIG.12

Szene Z0

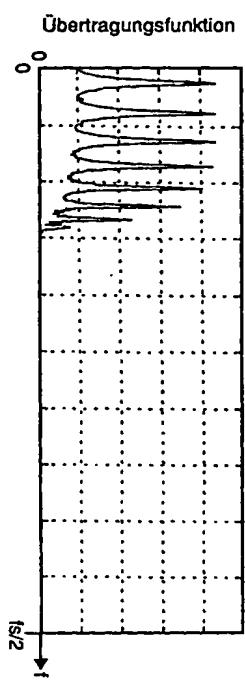
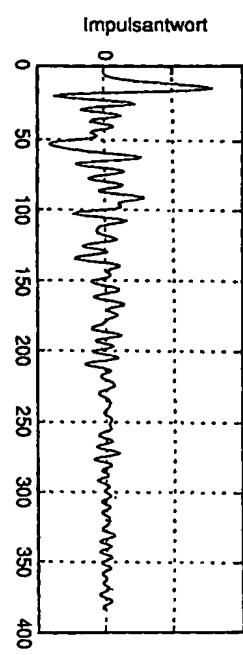


FIG.10

Szene Z1/1

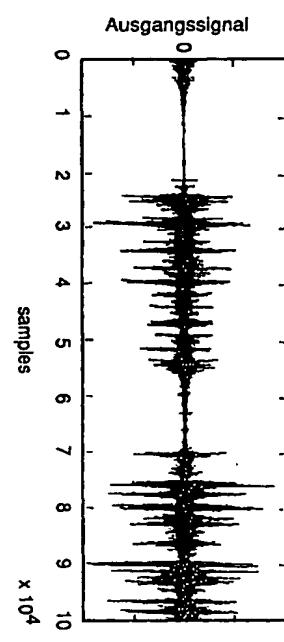
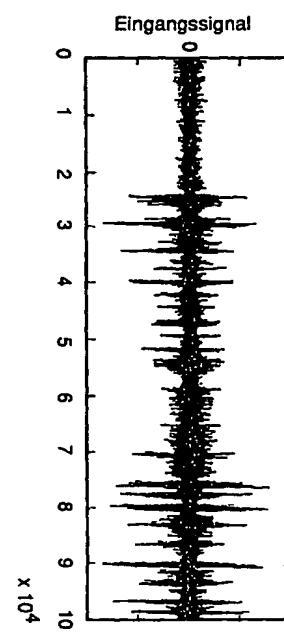


FIG.11

This Page is inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT OR DRAWING
- BLURED OR ILLEGIBLE TEXT OR DRAWING
- SKEWED/SLANTED IMAGES
- COLORED OR BLACK AND WHITE PHOTOGRAPHS
- GRAY SCALE DOCUMENTS
- LINES OR MARKS ON ORIGINAL DOCUMENT
- REPERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**  
**As rescanning documents *will not* correct images**  
**problems checked, please do not report the**  
**problems to the IFW Image Problem Mailbox**

**This Page Blank (uspto)**